

# Audio Book in Customized-voice (ABC)

2021 데이터 청년 캠퍼스



9조 박지현 서지혜 이승준 정상희 한유연



# CONTENTS



01

개발 배경

02

활용 데이터의 정의

03

데이터 처리방안/ 활용기법

04

개발 결과

05

서비스 활용 방안

06

개발효과

# 01

## 개발 배경 ABC가 필요한 이유



# 아이들을 위한 ABC

자연어의 감성분석 기술과 TTS 음성합성 기술을 융합한  
개인오디오북 제작 시스템

목소리를  
들려주세요!

사용자가 책을 읽는 자신의 목소리를 시스템에 들려주면,  
ABC는 사용자의 목소리에 더욱 실감나고 풍부한 감성을 담은 오디오북을 제공합니다.



# 아이들을 위한 customized 오디오북, 왜?



## 오디오북의 성장

기존 책 소비가 가장 많은  
3,40대 부터 멀티태스킹에 강한  
Z세대 까지 오디오북의 수요는  
꾸준히 증가



## 전자 기기 대중화

PC 태블릿이나 스마트폰의  
확산으로 인해 E-Book 활성화



## 코로나 장기화

코로나19가 장기화 되면서  
도서관과 같은 공공시설에  
이용 제한



## 맞벌이 부부 증가

맞벌이 부부 비율이 증가하면서  
육아 시간 감소

# 기존 오디오북과의 차이점

## "ABC의 강점"



- ✕ 성우들이 읽어주는 오디오북 존재
  - 제한된 수량으로 존재, 직접 녹음 된 것만 들을 수 있음
- ✕ AI가 읽어주는 오디오북 존재
  - 감정이 없는 모노톤의 목소리
- ✓ 사용자의 목소리를 모사해 생생하게 읽어주는 오디오북



아이들에게 반복해서 책을 읽어두지 않아도 OK  
 ⇒ 학습된 목소리로 어떤 동화책이든 여러 번 읽어줄 수 있음



감정을 다 살려서 읽지 않아도 OK  
 ⇒ 모노톤의 음성도 분위기에 맞는 실감나는 감정으로!

# 02

## 활용 데이터 정의



1 단계  
BERT

감성분석 단계:  
입력받은 동화책의 한 문장문장  
을 어떤 감정이 담겼는지 분석



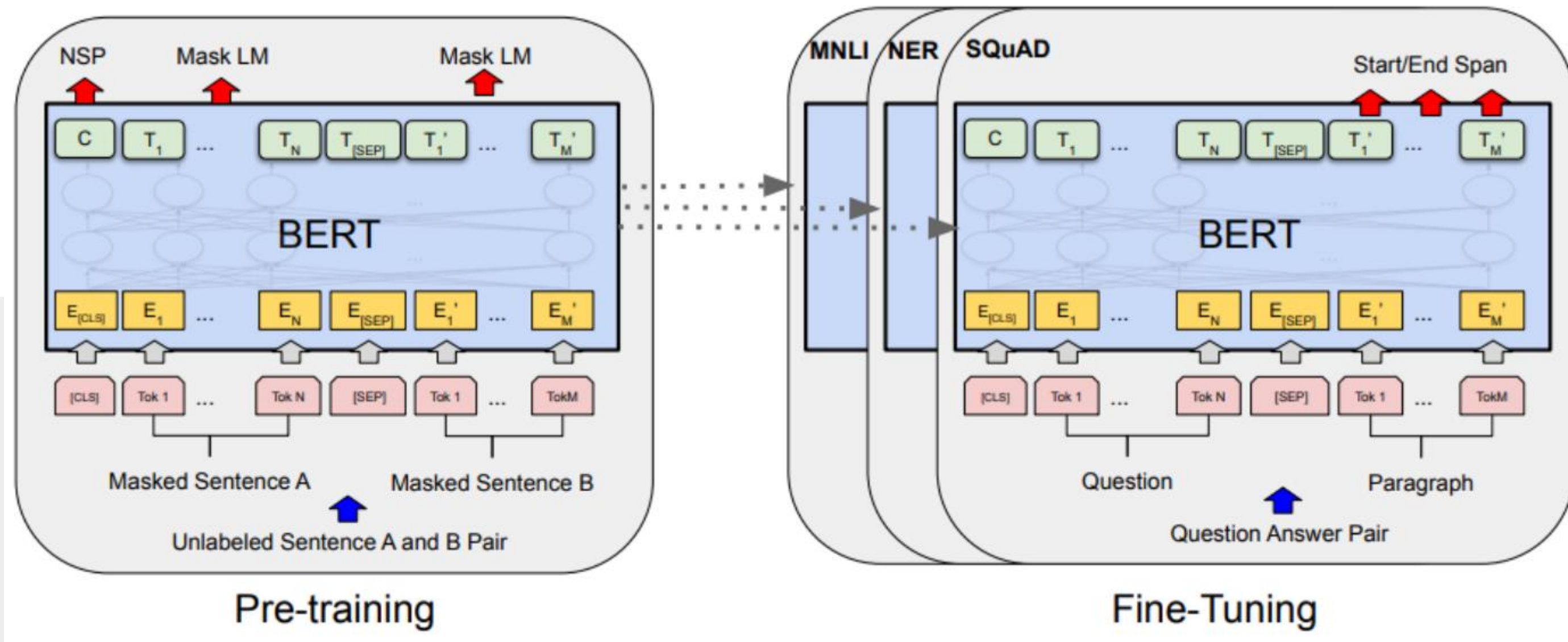
2 단계  
TACOTRON2 / Waveglow

음성합성 단계:  
해당 감정이 반영된 실감나는  
목소리로 동화책 읽어주기



# BERT

## 감성분석 단계



BERT - Transformer을 기반으로 한 사전 훈련 언어모델



# BERT

## 감성분석 단계 - 학습시킨 데이터

wav_id	발화문	상황	1번 감정	1번 감정세기	2번 감정	2번 감정세기	3번 감정	3번 감정세기	4번 감정	4번 감정세기	5번 감정	5번 감정세기
5ee1e7939aa8ea0eec53fac7	나 어제 헤어졌어	sad	Sadness	1	Sadness	1	Sadness	1	Sadness	1	Sadness	1
5ee1e7a379bf120ed2b81ba2	어쩌다 보니까 그렇게 됐네	sad	Sadness	1	Sadness	1	Neutral	0	Sadness	1	Sadness	1
5ed9793b9aa8ea0eec53f7f1	유기견 다큐멘터리를 봤는데 무책임한 사람들 때문에 너무 화가 나	disgust	Sadness	2	Sadness	2	Angry	1	Angry	1	Angry	1
5ed979582880d70f286128c3	버려지는 유기견들의 생활을 다룬 다큐멘터리 없어	disgust	Sadness	1	Sadness	2	Sadness	1	Angry	1	Sadness	2
5ed9797b1dcf350eeded50b8	작년보다 5배는 더 늘어나는 것 같대 최근 들어 더 늘어나고 있고	disgust	Fear	1	Sadness	1	Sadness	1	Angry	1	Sadness	2
5ed9799ac90a530ee56b5f6b	정보보조금이란 사람들의 기부금으로 운영되고 있어	disgust	Neutral	0	Neutral	0	Sadness	1	Sadness	1	Sadness	2
5ed979dc7e21a10eee253e90	맛이 벌어지는 대부분의 아이들이 다 큰 강아지 들었어 아직도 상처받	disgust	Sadness	2	Sadness	2	Sadness	2	Sadness	2	Sadness	2
5ed97a22c90a530ee56b5f6c	어제저녁에 진짜 무서웠어	fear	Fear	2	Fear	2	Fear	1	Fear	1	Fear	2
5ed97a381dcf350eeded50bc	친구랑 약속 있어 나갔는데 밥 먹고 이따가 지진이 발생하는 거야 얼	fear	Fear	2	Fear	2	Fear	2	Fear	1	Fear	1
5ed97a5cc90a530ee56b5f6d	큰 지진은 지나갔는데 여진이 조금씩 있는 거 같기도 해	fear	Fear	2	Fear	2	Fear	1	Fear	1	Fear	1
5ed97a779aa8ea0eec53f7f5	다른 테이블에서 밥 먹던 사람들이 도망치다가 넘어지고 그 와중에 나	fear	Sadness	1	Fear	2	Sadness	1	Fear	1	Fear	1
5ed97a977e21a10eee253e92	그렇지 않아도 오늘 친구 데리고 병원 가기로 했어	fear	Neutral	0	Sadness	1	Sadness	1	Sadness	1	Fear	2
5ed97ac29aa8ea0eec53f7f8	룸메이트와 너무 자주 싸우게 돼	anger	Disgust	2	Angry	2	Angry	1	Angry	1	Fear	1

1. AI HUB의 감정 분류를 위한  
대화 음성 데이터셋  
(텍스트 데이터만 사용)

2. 카이스트 오디오북 데이터셋  
(텍스트 데이터만 사용)

	A	B
1	Index	대사
2	1	다람쥐와 호랑이
3	2	추운 겨울날, 배고픈 호랑이가 작은 다람쥐를 잡았어요.
4	3	"어휴, 배고픈데 잘됐네.
5	4	너를 잡아먹으면 되겠군."
6	5	"잠깐만요!
7	6	호랑이님, 저 같은 작은 다람쥐만으로는 부족하지 않으신가요?
8	7	저를 살려주시면 제가 물고기를 잔뜩 먹게 해 드릴게요."
9	8	다람쥐의 말에 호랑이는 귀가 솔깃했어요.
10	9	"뭐라고?
11	10	물고기를 잔뜩 먹게 해준단 말이야?"
12	11	"네.
13	12	제가 누구예요?
14	13	꽤 많은 다람쥐잖아요.
15	14	저만 따라오세요."
16	15	토끼는 호랑이를 연못으로 데려갔죠.
17	16	"호랑이님의 긴 꼬리를, 자, 자, 이렇게.
18	17	이렇게.
19	18	물 속에 넣고 있으면 물고기들이 꼬리를 물 거예요."
20	19	"그래?
21	20	정말이지?"

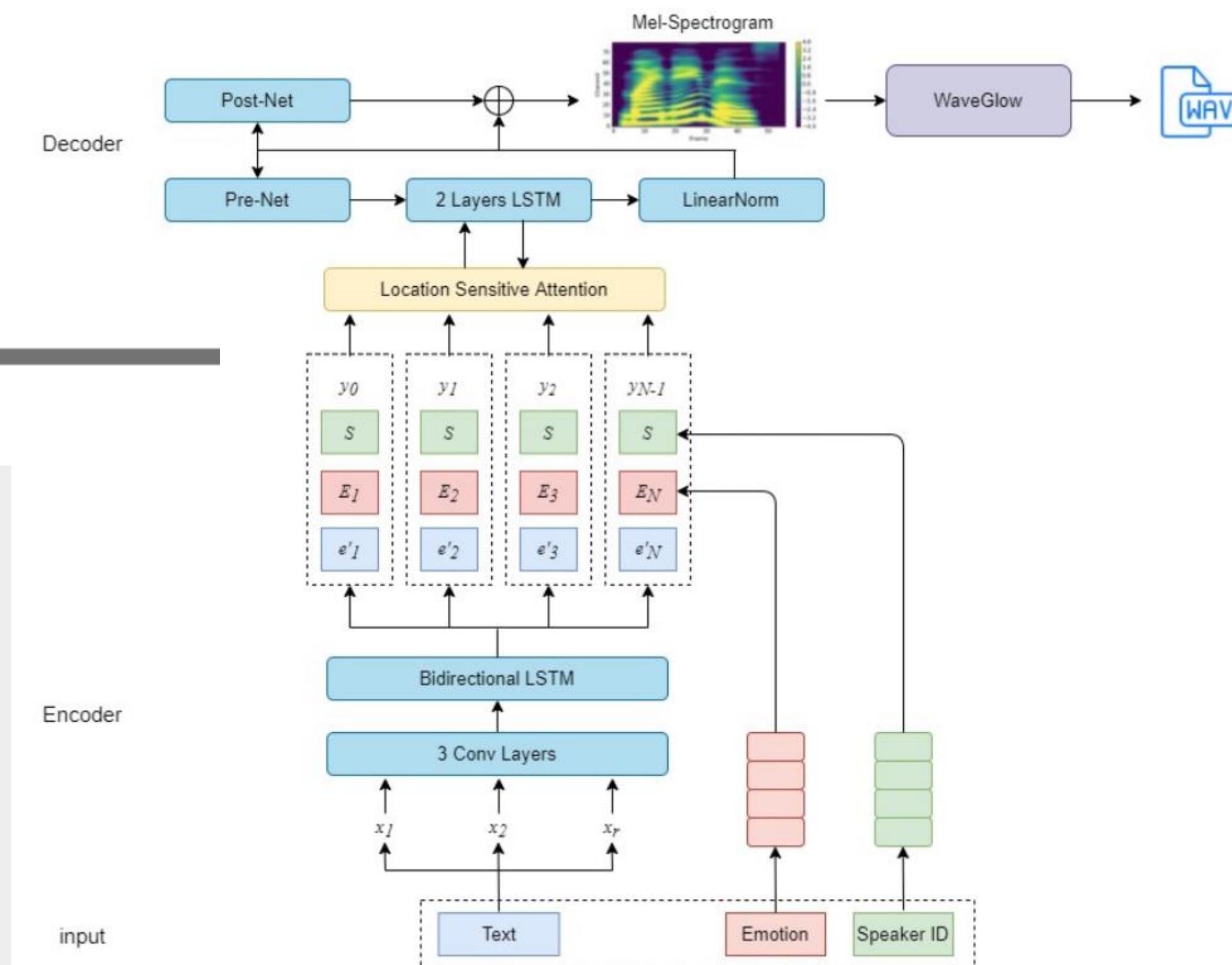
# 타코트론2+ Waveglow

## 음성합성 단계

TACOTRON2 : Attention 기반 Seq-to-Seq 모델

- Encoder : 텍스트를 일련 길이의 벡터로 변환
- Decoder : Attention에서 얻은 정보를 이용하여 Emotion이 담긴 mel-spectrogram을 생성

Vocoder : mel-spectrogram을 오디오파일인 wav 로 변환하는 역할. Waveglow를 Vocoder로 사용



# 타코트론2+ Waveglow

## 음성합성 단계 - 학습시킨 데이터

### 1. Selvas AI 감정 화자 데이터 -> 타코트론2 / Waveglow 학습

- 감정 대본을 사용하여 여성 5인, 남성 5인이 녹음
- 4가지 감정: 일반, 기쁨, 화남, 슬픔
- 400 문장 \* 10명

##텍스트 파일 예시

```
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_1_L.wav|과학적 지식은 어떻게 생성될까?|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_1_R.wav|과학적 지식은 어떻게 생성될까?|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_2_L.wav|이에 대한 설명은 과학 철학적 관점에 따라 달라질 수 있다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_2_R.wav|이에 대한 설명은 과학 철학적 관점에 따라 달라질 수 있다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_3_L.wav|그 중 하나가 경험적 검증 가능성에 의해 과학적 진술의 의미를 평가하는
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_3_R.wav|그 중 하나가 경험적 검증 가능성에 의해 과학적 진술의 의미를 평가하는
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_4_L.wav|언어의 회귀에 대한 연구 과정을 통해 과학적 지식의 생성 과정을 논리 실증
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_4_R.wav|언어의 회귀에 대한 연구 과정을 통해 과학적 지식의 생성 과정을 논리 실증
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_5_L.wav|과학자들은 언어가 어떻게 태어난 곳으로 돌아오는지 알고 싶었다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_5_R.wav|과학자들은 언어가 어떻게 태어난 곳으로 돌아오는지 알고 싶었다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_6_L.wav|인디언들은 초자연적인 힘에 의해 언어가 회귀한다고 믿고 있었는데, 과학자
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_6_R.wav|인디언들은 초자연적인 힘에 의해 언어가 회귀한다고 믿고 있었는데, 과학자
```

### 2. 조원 박지현 음성 데이터 -> 타코트론2 추가 학습

- 음성 모사 시 transfer 학습을 위해 사용
- 한 문장씩 나뉜 음성 데이터, 약 1시간 10분 분량
- 화자의 음성은 최대한 monotonous하게 녹음
- 타코트론2 모델의 학습을 input으로 쓰기 위해  
다음과 같이 텍스트 파일 저장



# 데이터 처리 방안/ 활용 기법

BERT/TACOTRON2/waveglow





# BERT

## 데이터 처리 방안

Index	대사
1	다람쥐와 호랑이
2	추운 겨울날, 배고픈 호랑이가 작은 다람쥐를 잡았어요.
3	"어흥, 배고픈데 잘됐네.
4	너를 잡아먹으면 되겠군."
5	"잠깐만요!
6	호랑이님, 저 같은 작은 다람쥐만으로는 부족하지 않으신가요?
7	저를 살려주시면 제가 물고기를 잔뜩 먹게 해 드릴게요."
8	다람쥐의 말에 호랑이는 귀가 솔깃했어요.
9	"뭐라고?
10	물고기를 잔뜩 먹게 해준단 말이야?"
11	"네.
12	제가 누구예요?
13	꽤 많은 다람쥐잖아요.
14	저만 따라오세요."
15	토끼는 호랑이를 연못으로 데려갔죠.
16	"호랑이님의 긴 꼬리를, 자, 자, 이렇게.
17	이렇게.
18	물 속에 넣고 있으면 물고기들이 꼬리를 물 거예요."
19	"그래?
20	정말이지?"
21	"아, 그럼요!
22	물고기들이 잔뜩 매달려서 꼬리가 묵직해지면 얼른 꼬리를 뺏아채기만 하면 되요."
23	호랑이는 다람쥐가 시키는 대로



Index	대사		
1	다람쥐와 호랑이	neutral	2
2	추운 겨울날, 배고픈 호랑이가 작은 다람쥐를 잡았어요.	neutral	2
3	"어흥, 배고픈데 잘됐네.	neutral	2
4	너를 잡아먹으면 되겠군."	neutral	2
5	"잠깐만요!	sad	1
6	호랑이님, 저 같은 작은 다람쥐만으로는 부족하지 않으신가요?	fear	2
7	저를 살려주시면 제가 물고기를 잔뜩 먹게 해 드릴게요."	happy	1
8	다람쥐의 말에 호랑이는 귀가 솔깃했어요.	happy	1
9	"뭐라고?	neutral	2
10	물고기를 잔뜩 먹게 해준단 말이야?"	neutral	2
11	"네.	neutral	2
12	제가 누구예요?	happy	2
13	꽤 많은 다람쥐잖아요.	happy	2
14	저만 따라오세요."	happy	1
15	토끼는 호랑이를 연못으로 데려갔죠.	neutral	2
16	"호랑이님의 긴 꼬리를, 자, 자, 이렇게.	neutral	2
17	이렇게.	neutral	2
18	물 속에 넣고 있으면 물고기들이 꼬리를 물 거예요."	neutral	2
19	"그래?	neutral	2
20	정말이지?"	happy	1
21	"아, 그럼요!	happy	2
22	물고기들이 잔뜩 매달려서 꼬리가 묵직해지면 얼른 꼬리를 뺏아채기만 하면 되요."	happy	1
23	호랑이는 다람쥐가 시키는 대로	neutral	2
24	꼬리를 물 속 깊이 댔어요	neutral	2

카이스트 오디오북 데이터셋의 텍스트 데이터  
- 텍스트 형식의 동화책 문장을 문장마다 감정태깅

# TACOTRON2

## 데이터 처리 방안

조원 박지현 음성 데이터

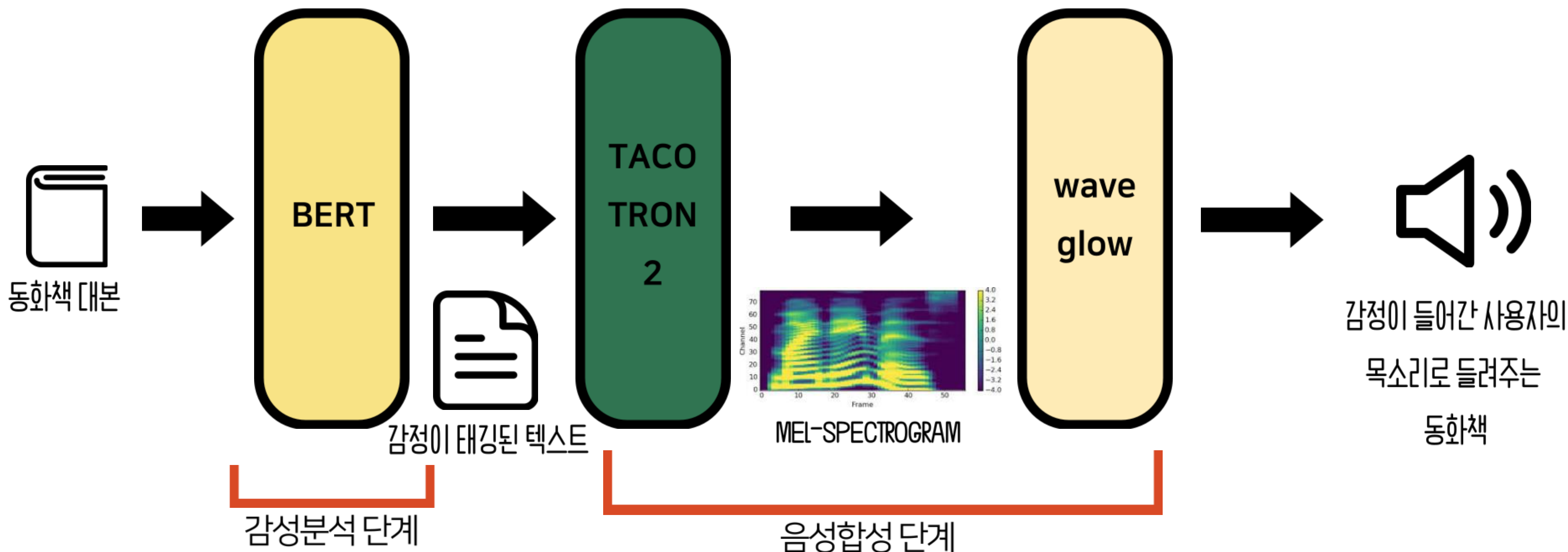
- Adobe Audition 프로그램을 사용해 데이터 처리
- Sample Rate : 22050Hz
- Channel : Mono
- 1시간 4분의 분량을 문장 별로 분리
- 'wav파일경로|텍스트|speaker\_id|감정|감정' 형식으로 텍스트 파일을 저장함

##텍스트 파일 예시

```
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_1_L.wav|과학적 지식은 어떻게 생성될까?|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_1_R.wav|과학적 지식은 어떻게 생성될까?|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_2_L.wav|이에 대한 설명은 과학 철학적 관점에 따라 달라질 수 있다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_2_R.wav|이에 대한 설명은 과학 철학적 관점에 따라 달라질 수 있다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_3_L.wav|그 중 하나가 경험적 검증 가능성에 의해 과학적 진술의 의미를 판가름하는
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_3_R.wav|그 중 하나가 경험적 검증 가능성에 의해 과학적 진술의 의미를 판가름하는
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_4_L.wav|언어의 회귀에 대한 연구 과정을 통해 과학적 지식의 생성 과정을 논리 실증
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_4_R.wav|언어의 회귀에 대한 연구 과정을 통해 과학적 지식의 생성 과정을 논리 실증
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_5_L.wav|과학자들은 언어가 어떻게 태어난 곳으로 돌아오는지 알고 싶었다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_5_R.wav|과학자들은 언어가 어떻게 태어난 곳으로 돌아오는지 알고 싶었다.|pjh|0|0
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_6_L.wav|인디언들은 초자연적인 힘에 의해 언어가 회귀한다고 믿고 있었는데, 과학자
/content/drive/MyDrive/emotion-tts/jvoice/jvoice_6_R.wav|인디언들은 초자연적인 힘에 의해 언어가 회귀한다고 믿고 있었는데, 과학자
```

# 전체적인 프로세스

BERT/ 타코트론2/ vocoder(waveglow)





# 04

## 개발 결과



# 결과 비교

## AI 목소리와 학습시킨 조원 목소리랑 비교



기존 화자의 목소리



새로운 화자의 목소리(지현님)

“예나 예나에 모두의 사랑을 받는 작고 귀여운 소녀가 있습니다.  
하지만 그 소녀를 가장 사랑하는 것은 그녀의 할머니였습니다..”

05

## 서비스 활용 방안

### 어플리케이션으로 사용

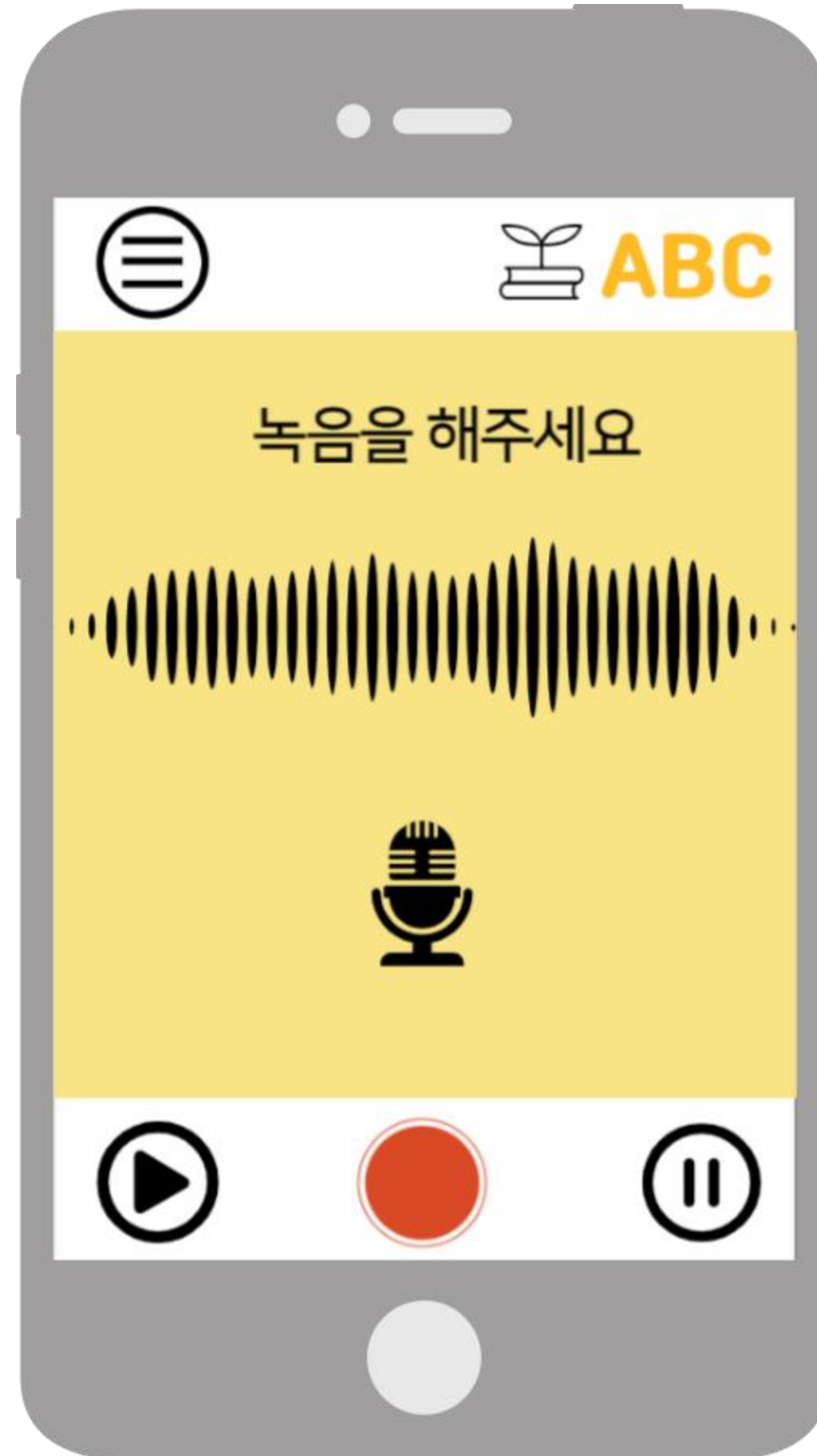


## STEP 1.

사용자의 목소리 학습을 위해  
한시간 분량을 책을 읽어 주세요.

## Step 2.

텍스트 형태의 동화책을 입력해주세요.



## STEP 3.

조금만 기다려주세요.

## STEP 4.

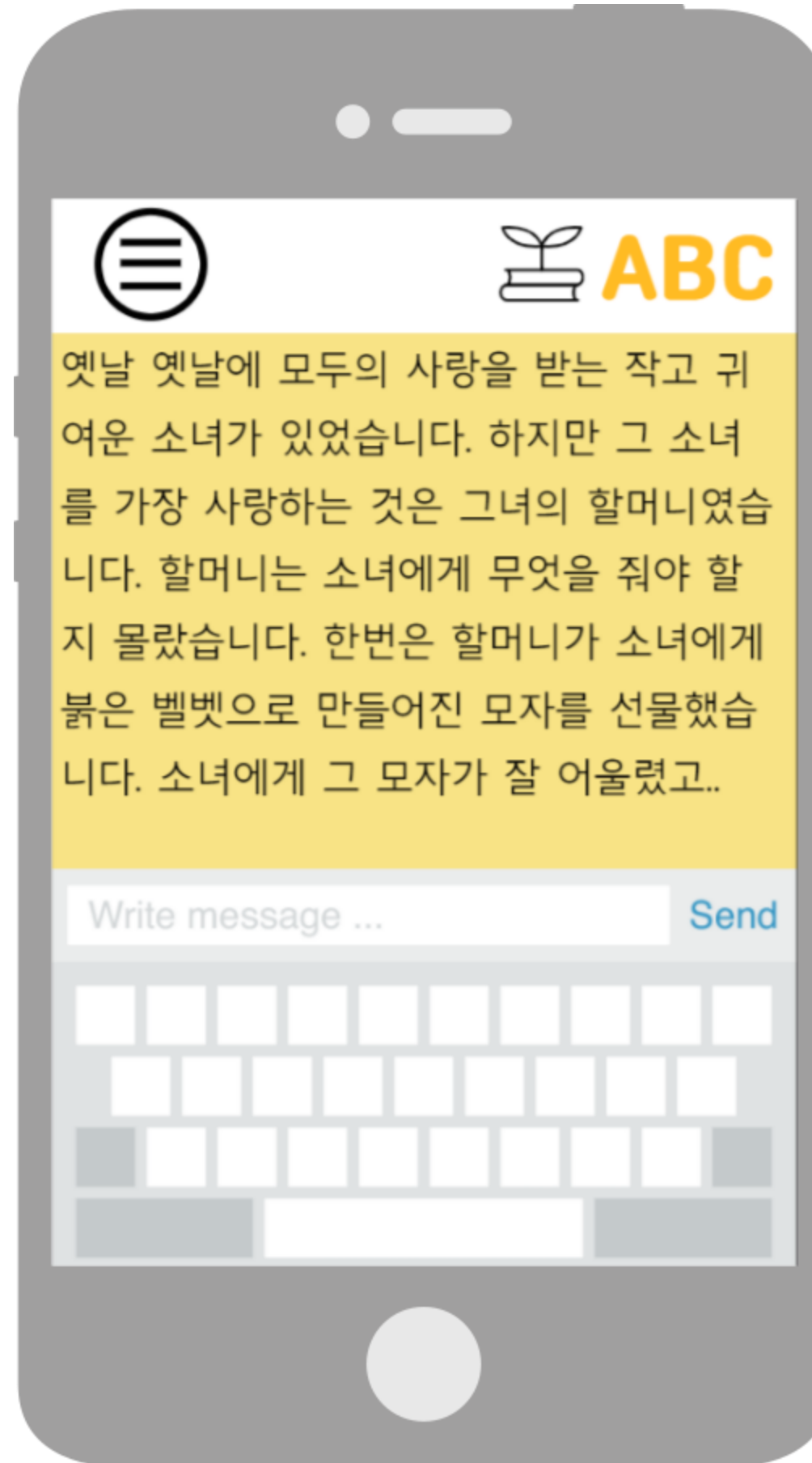
이제 아이에게 동화책을 들려주세요!

## Step 1.

사용자의 목소리 학습을 위해  
한시간 분량을 책을 읽어 주세요.

## STEP 2.

텍스트 형태의 동화책을  
입력해주세요.



## STEP 3.

조금만 기다려주세요.

## STEP 4.

이제 아이에게 동화책을 들려주세요!

## Step 1.

사용자의 목소리 학습을 위해  
한시간 분량을 책을 읽어 주세요.

## Step 2.

텍스트 형태의 동화책을 입력해주세요.



## STEP 3.

조금만 기다려주세요..!

## STEP 4.

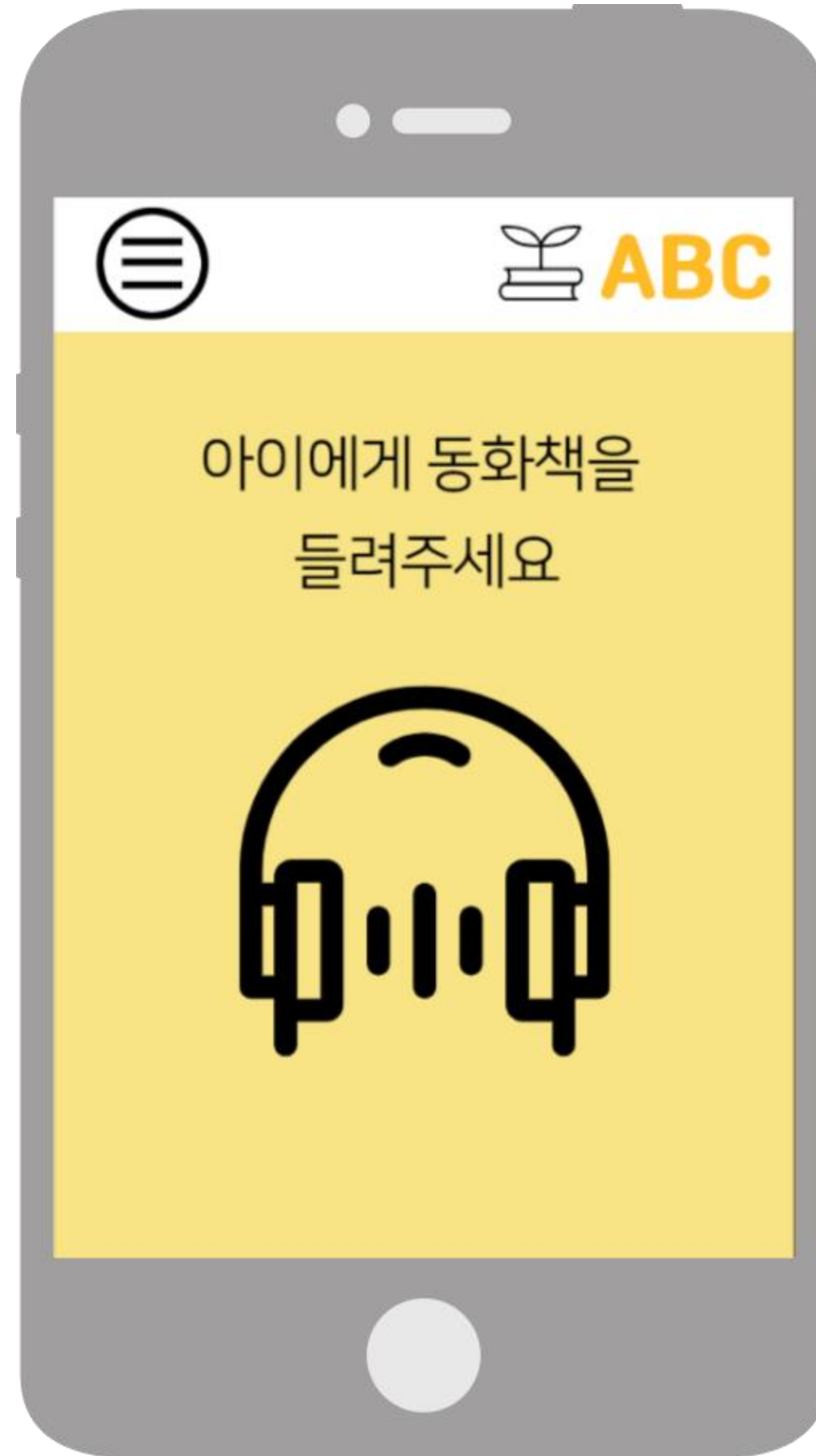
이제 아이에게 동화책을 들려주세요!

## Step 1.

사용자의 목소리 학습을 위해  
한시간 분량을 책을 읽어 주세요.

## Step 2.

텍스트 형태의 동화책을 입력해주세요.



## STEP 3.

조금만 기다려주세요.

## STEP 4.

이제 아이에게 동화책을 들려주세요!





# 기대효과

## 4가지 측면

사회적 효과



경제적 효과



교육적 효과



시장의 확장



## 사회적 효과

돌봄의 손실이 필요한 아이들을 대상으로 하는 사회적 서비스

육아에 투자할 시간적 여유가 없는 부모 (맞벌이 부부)  
낮- 저녁 시간대에 홀로 남겨지는 아이

} 공동체(가정)의 복리

그 밖에도

- 저소득층
- 한부모 가정
- 고아원

} 소외계층의 복리

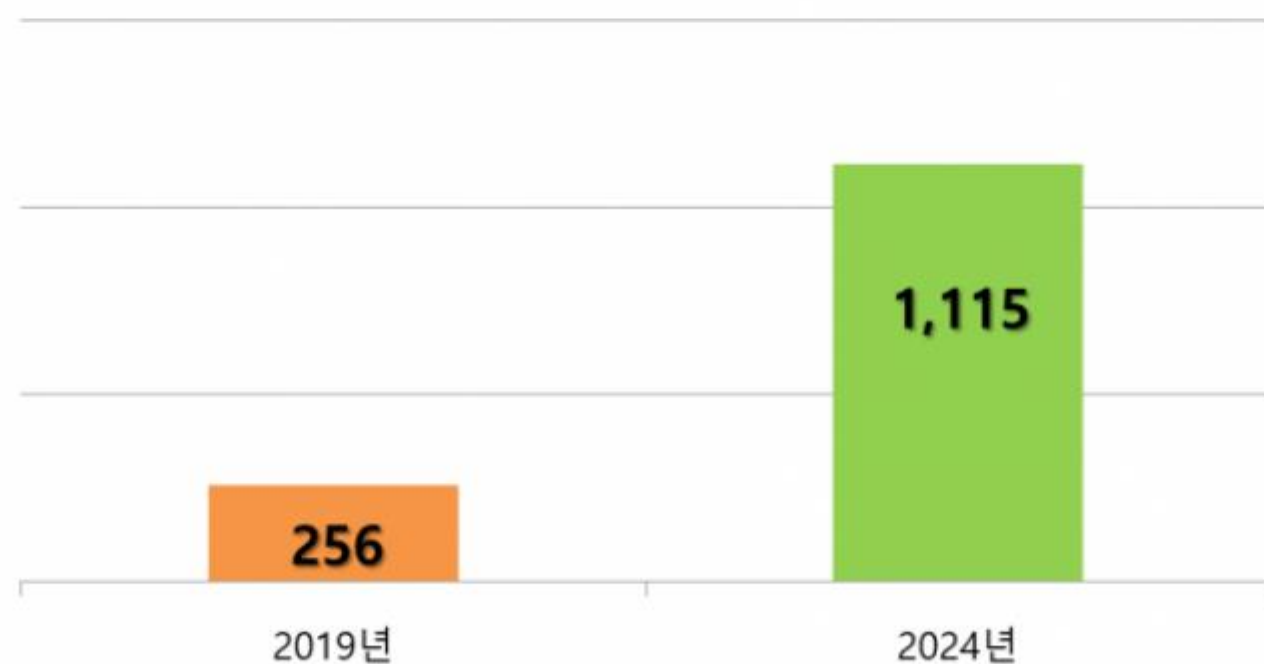


맞벌이 부부 비율 ↑

## 경제적 효과

규모 ↑, 접근성 ↑ 오디오북 시장에서의  
이윤 창출 서비스

국내 오디오 콘텐츠 시장 성장 추정  
(단위: 억원)



해마다 인기와 성장을 거듭하는 오디오 콘텐츠 분야

"읽지말고 편히 들으세요"...오디오북 전성시대

[트렌드 취재] 오디오 시장의 화려한 부활

걷고, 먹고, 자는 순간까지 함께하는 '듣는' 책의 매력

'편리미엄' 트렌드에 쑥쑥 크는 오디오북 시장

## 교육적 효과

아이들의 흥미를 이끌어내는 교육 서비스

- 시간과 장소에 구애 받지 않는다 ✓
- 간편하게 전자기기로 듣는다 ✓
- "부모님의 목소리"가 담긴 동화책 ✓

실감나는 감정표현과 목소리의 모사를 통한 흥미 유발  
→ 독서 습관 정립



독서 습관 정립 및 정서적 안정

# 교육적 효과

아이들의 흥미를 이끌어내는 교육 서비스

부모의 반복된 책 읽어주기가 유아의  
독자적 읽기에 미치는 영향

Effects of Parents' Repeated Reading of a Storybook  
on Young Children's Independent Reading

현 은 자\*  
Hyun, Eun Ja

## ABSTRACT

The purpose of this study was to examine the effect of parents' repeated reading of the same storybook on young children's independent reading. The subjects were forty five- and six-year-old children and their parents. The children attended four preschools and kindergartens in three cities, Ann Arbor, Northvil-

## 생애 초기 발달이 결정적...책읽어주기 효과

### ◆ 책 읽어주기 독서활동, 인지능력 발달에 도움

삼각산초는 다양한 형태의 책 읽어주기 활동을 펼쳐가고 있다. 우선, 교장이 특정을 정해 아침 시간에 전교생을 대상으로 책을 직접 읽어준다. 그 시간만큼은 아이의 모든 시선이 책에 쏠리게 된다. 그러면 독서교육 하나만으로도 시각적·청각적 인지가 많이 발달한다는 것이란 것이 사실 교장의 선행이다.



시장 확장

## 시장의 확장

타겟 시장 -  
B2C에서 B2G로의 확장

개인(부모님)에서 교육기관(유치원/초등학교)으로  
시장을 확장

# Thank you

덕조 박지현 서지혜 이승준 정상희 한유연

